

Estatística Básica I

INTRODUÇÃO À ESTATÍSTICA

TUANY CASTRO

O que é estatística?

A estatística pode ser entendida como um conjunto de técnicas que permite, de forma sistemática, organizar, descrever, analisar e interpretar dados oriundos de estudos ou experimentos, realizados em qualquer área do conhecimento.

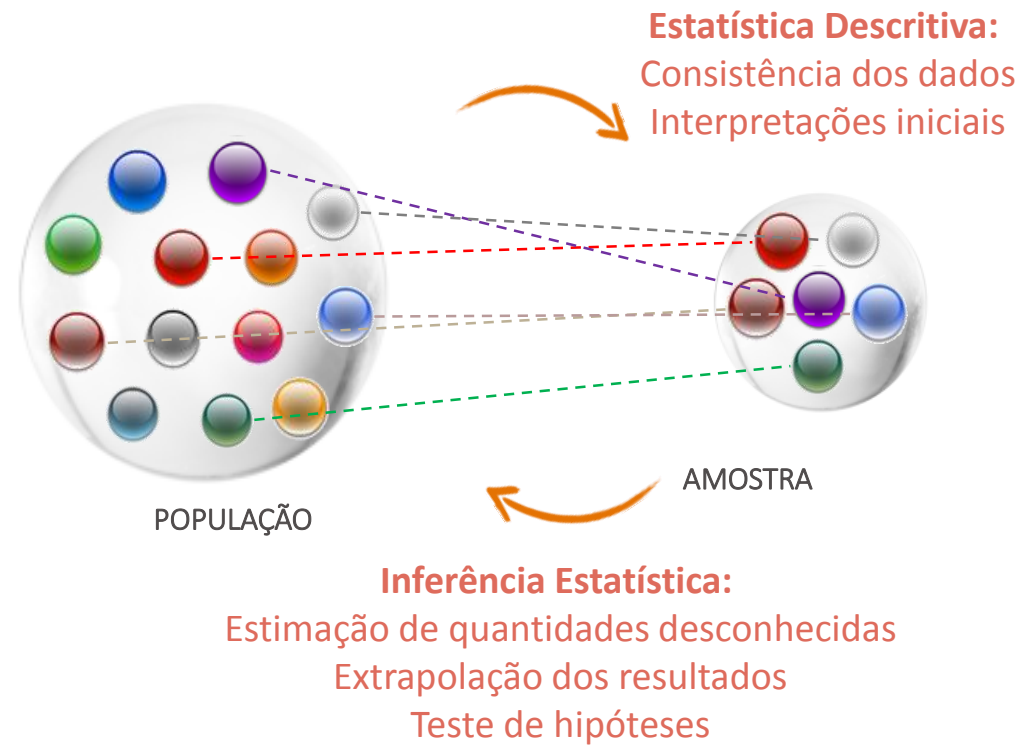


Áreas da estatística

☐ Estatística Descritiva

☐ Probabilidade

☐ Inferência Estatística



Etapas desejáveis para boa análise estatística

- ❑ Definição (operacional) do problema
- ❑ Planejamento do estudo
- ❑ Coleta e armazenamento de dados
- ❑ Análise Descritiva/Correção do conjunto de dados
- ❑ Análise Inferencial
 - Modelo Probabilístico
 - Tradução dos objetivos em termos dos elementos do modelo probabilístico
 - Ajuste do modelo estatístico
 - Avaliação dos resultados
- ❑ Tradução dos resultados para aplicação prática

Planilha de dados

- ☐ Não utilizar limitadores de celas (borders) ou cores;
- ☐ Reservar apenas primeira linha para rótulos das variáveis;
- ☐ Não usar acentos ou outros símbolos nos rótulos;
- ☐ Não esquecer uma coluna para a variável indicadora das unidades de investigação (evitar informações confidenciais como nomes de pacientes);
- ☐ Escolher ponto ou vírgula para separação de casas decimais;
- ☐ Especificar o número de casas decimais;
- ☐ Incluir dicionário.

Exemplo de planilha

ident	idade	nummal	parasita	numgest	idgest	sexorn	pesorn	estrn
1	25	0	0	3	38	2	3665	46
2	30	0	0	9	37	1	2880	44
3	40	0	2	1	41	1	2960	52
4	26	3	0	2	40	1	2740	47
5	.	0	0	1	38	1	2975	50
6	18	0	0	.	38	2	2770	48
7	20	0	0	1	41	1	2755	48
8	15	0	0	1	39	1	2860	49
9	.	0	0	.	42	2	3000	50
10	18	0	4	1	40	1	3515	51
11	17	2	0	2	40	1	3645	54
12	18	1	1	3	40	2	2665	48
13	30	0	0	6	40	2	2995	49
14	19	0	0	1	40	1	2972	46
15	32	0	0	5	41	2	3045	50
34	.	0	0	.	39	2	2950	47

Dicionário

Rótulos	Variável	Unidade de medida
idade	Idade da mãe	anos
nummal	Quantidade de malárias durante a gestação	número inteiro
parasita	Espécie do parasita da malária	0: não infectada
		1: P. vivax
		2: P. falciparum
		3: malária mista
		4: indeterminado
numgest	Paridade (quantidade de gestações)	Número inteiro
idgest	Idade gestacional no parto	semanas
sexorn	Sexo do recém-nascido	1: masculino
		2: feminino
pesorn	Peso do recém-nascido	g
estrn	Estatura do recém-nascido	cm
pcefal	Perímetro cefálico do recém-nascido	cm
Obs:	Observações omissas são representadas por um ponto	

Pesadelo

nome	idade	local	Data US	Cir-retosigmoide	US - retosigmoide	US-retosigmoide-LONG	Preparo US
Anne	23	MSA	16/04/2003	não	não		
Aparecida		HC	17.04.01 (verificar se tem US 08.2001)	não??	sim.Reto 3,3x1,5 cm	3,3	
Arlene		HC (Mauricio-Santa Joana)	24/07/2002	Verificar.	não		bom
Assunta	52	HC	22/07/2003	não	não		não
Beatriz		MSA	02/04/2004	não	não		bom
Bernadete	40	HC	15/04/2003	sim	sim.	2,6	bom
Bettina		MSA	26/05/2004	não	não		bom
Blanca		HC	23/07/2003	não	não		regular
Carla		MSA	23/10/2002	sim	não		
Carla		MSA	06/02/2004	sim	Sim,	4,5	
Carolina	24	MSA	25/07/2003	não	sim. Sigmoide: 1,0x0,4cm.	10	
Cassia kolaya		MSA	04/06/2004	sim	Sim 3,5x1,0x2,1	3,5	bom

Estatística Descritiva

A Estatística Descritiva é um conjunto de técnicas destinadas a descrever e resumir os dados, a fim de que possamos tirar conclusões a respeito de características de interesse.

Estatística Descritiva

Como “tratar” um conjunto de dados para extrair informações?

- ☐ Tabelas de frequência
- ☐ Gráficos
- ☐ Medidas resumo



Exemplo

Suponha que um questionário aplicado aos alunos de uma escola forneça as seguintes informações:

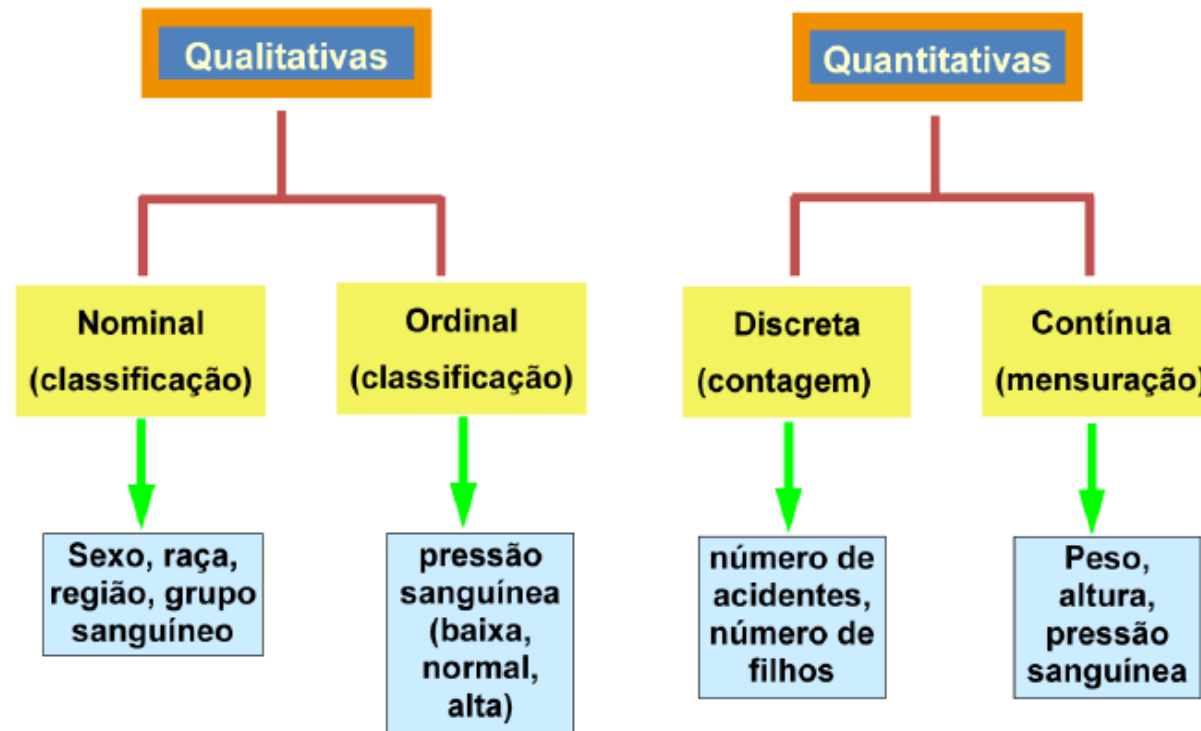
- ☐ Id: identificação do aluno
- ☐ Turma: turma a que o aluno foi alocado
- ☐ Sexo: F se feminino, M se masculino
- ☐ Idade: idade em anos
- ☐ Alt: altura em metros
- ☐ Peso: peso em quilogramas
- ☐ Filhos: número de filhos na família
- ☐ Fuma: hábito de fumar, sim ou não
- ☐ Toler: tolerância ao cigarro (I se indiferente, P se incomoda pouco e M se incomoda muito)
- ☐ Exer: horas de atividade física, por semana

Exemplo

Id	Turma	Sexo	Idade	Alt	Peso	Filhos	Fuma	Toler	Exer
1	A	F	17	1,6	60,5	2	Não	P	0
2	A	F	18	1,7	55,0	1	Não	M	0
3	A	M	20	1,9	72,8	2	Não	P	5
4	A	M	25	1,9	80,9	2	Não	P	5
5	A	F	19	1,6	49,0	1	Não	M	2
6	A	M	18	1,7	73,0	3	Não	M	2
7	A	F	17	1,6	58,0	1	Não	P	3
8	A	F	19	1,7	55,0	3	Sim	I	2
9	A	F	18	1,5	44,0	2	Não	M	3
10	A	F	17	1,6	49,0	1	Não	P	2
11	A	M	18	1,7	73,0	2	Sim	I	2
12	B	F	17	1,7	52,5	1	Não	P	10
13	B	F	17	1,6	54,5	1	Não	P	6
14	B	F	21	1,6	49,0	3	Sim	I	5
15	B	M	18	1,7	58,0	1	Não	M	4
16	B	M	18	1,7	87,0	2	Não	M	0
17	B	M	25	1,8	75,0	1	Não	P	3
18	B	M	20	1,7	73,0	3	Não	M	3
19	B	F	20	1,7	56,0	2	Sim	I	1
20	B	F	18	1,7	58,0	1	Não	P	4

Classificação de variáveis

O tratamento dos dados depende do tipo da variável.



Variáveis quantitativas

Características descritas como números.

- ❑ **Discretas:** resultantes de contagens, assumindo assim, valores inteiros.
- ❑ **Contínuas:** assumem valores em intervalos e, geralmente, provenientes de mensuração.

No exemplo...

Variáveis: Turma, Sexo, Idade, Altura, Peso, Filhos, Fuma, Tolerância e Exercício

Quantitativas Discretas

Filhos

Quantitativas Contínuas

Idade
Altura
Peso
Exercício

Variáveis qualitativas

Características representam atributos.

- ❑ **Nominais:** não existe ordenação entre as categorias.
- ❑ **Ordinais:** existe ordenação, indicando intensidades crescentes de realização.

No exemplo...

Variáveis: Turma, Sexo, Idade, Altura, Peso, Filhos, Fuma, Tolerância e Exercício

Qualitativas Nominais

Turma
Sexo
Fuma

Qualitativas Ordinais

Tolerância

Exercício

- **Sexo:** qualitativa nominal
- **Número de Filhos:** quantitativa discreta
- **Cor dos olhos:** qualitativa nominal
- **Classe Social:** qualitativa ordinal
- **Peso:** quantitativa contínua
- **Número de carros em um estacionamento:** quantitativa discreta
- **Número de chamadas recebidas por uma central telefônica:** quantitativa discreta



Exercício

- **Grau de Instrução:** qualitativa ordinal
- **Peça defeituosa (sim ou não):** qualitativa nominal
- **Idade:** quantitativa contínua
- **Número de passageiros de um voo:** quantitativa discreta
- **Qualidade de um produto (ruim, regular, bom):** qualitativa ordinal
- **Altura:** quantitativa contínua
- **Salário:** quantitativa contínua



Observações

Em muitas situações práticas, a classificação das variáveis depende de certas particularidades, de maneira que podemos *discretizar* uma variável contínua ou mesmo coletar uma variável quantitativa como qualitativa.

Exemplo 1: A variável Idade, medida em anos, é quantitativa contínua, porém, se coletada em faixas etárias, é qualitativa ordinal.

Exemplo 2: A variável Idade, medida em número de anos, pode ainda ser vista como discreta para obtermos uma melhor representação da ocorrência de seus valores.
